Acta Applicandae Mathematicae **00:** 1–13, 2003.

A New Estimator for a Tail Index

V. PAULAUSKAS
Department of Mathematics and Informatics, Vilnius University and Institue of Mathematics and Informatics, Akademijos 4, LT-2021 Vilnius, Lithuania

(Received: 31 August 2002)

Abstract. We investigate properties of a new estimator for a tail index introduced by Davydov and co-workers. The main advantage of this estimator is the simplicity of the statistic used for the estimator. We provide results of simulation by comparing plots of our's and Hill's estimators.

Mathematics Subject Classifications (1991):

Key words:

1. Introduction and Formulation of Results

During past few decades, in many fields of applied probability, more and more attention has been paid to heavy-tailed distributions and, as a consequence, the problem of estimation of a tail index from various types of data has become rather important. At present, there are several known estimators, all of them expressed as some functionals of order statistics of a sample, the best known among them being the one proposed by Hill in [15].

Let us consider a sample of size n taken from a heavy-tailed distribution function F, that is, we assume that $X_1, \dots, X_n$ are independent identically distributed (i.i.d.) random variables with a distribution function F satisfying the following relation for large x:

$$1 - F(x) = x^{-\alpha} L(x), \tag{1}$$

where $\alpha > 0$, and L is slowly varying:

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1.$$

Let $X_{n,1} \leqslant X_{n,2} \leqslant \cdots \leqslant X_{n,n}$ denote the order staistics of $X_1, \dots, X_n$. The following estimator to estimate the parameter $\gamma = 1/\alpha$ was proposed in [15]:

$$\gamma_{n,k}^{(1)} = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n,n-i} - \log X_{n,n-k},$$

where k is some number satisfying $1 \leqslant k \leqslant n$. The problem of how to choose k is rather complicated (see, for example, [4, 10, 11, 13, 14, 16]). During the 25 years after this estimator, which bears Hill's name, was proposed there have been many papers devoted to the problem of choosing k and some properties and modifications of Hill's estimator, see, for example, [2, 3, 8, 9, 19] and references therein. Here are several other estimators of the parameter γ, based on order statistics:

$$\gamma_{n,k}^{(2)} = (\log 2)^{-1} \log \frac{X_{n,n-[k/4]} - X_{n,n-[k/2]}}{X_{n,n-[k/2]} - X_{n,n-k}},$$
$$\gamma_{n,k}^{(3)} = \gamma_{n,k}^{(1)} + 1 - \tfrac{1}{2}\big(1 - (\gamma_{n,k}^{(1)})^2/M_n\big)^{-1},$$
$$\gamma_{n,k}^{(4)} = \frac{M_n}{2\gamma_{n,k}^{(1)}},$$

where

$$M_n = \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n,n-i} - \log X_{n,n-k})^2.$$

The estimator $\gamma_{n,k}^{(2)}$ was proposed by Pickands in [19], $\gamma_{n,k}^{(3)}$ – in [9] and $\gamma_{n,k}^{(4)}$ – by C. G. de Vries [7]. We separated and presented these four estimators because in [7], all these estimators are compared and it is shown that none of the estimators dominates the others. It turned out that for different values of the parameters γ and ρ (the parameters characterizing the so-called second-order asymptotic behavior of F, which will be introduced below) different estimators have the smallest asymptotic mean-squared error. As was mentioned, all these estimators are based on order statistics, therefore investigation of their asymptotic properties is not a simple task. In [5] (see also [6]), we proposed a new estimator, based on a different idea (which, in turn, was born through the consideration rather abstract objects – random compact convex sets) and proved a simple result – the asymptotic normality of this estimator.

Here we investigate in more detail this new estimator and one of the purposes of the paper is to draw the attention of statisticians to this, since preliminary simulation results are rather promising.

Let us assume that we have a sample $X_1, \ldots, X_N$ from distribution F, which satisfies the second-order asymptotic relation (as $x \to \infty$)

$$1 - F(x) = C_1 x^{-\alpha} + C_2 x^{-\beta} + \mathrm{o}(x^{-\beta}), \tag{2}$$

with some parameters $0 < \alpha < \beta \leqslant \infty$. The case $\beta = \infty$ corresponds to Pareto distribution, $\beta = 2\alpha$ – to stable distribution with exponent $0 < \alpha < 2$. (In [7] the second-order asymptotic relation is used in a different form with parameters $\gamma = 1/\alpha$ and ρ, and there is a simple relation between (α, β) and (γ, ρ).)

We divide the sample into n groups $V_1, \ldots, V_n$, each group containing m random variables, that is, we assume that $N = n \cdot m$. (In practice, we choose m and then $n = [N/m]$, where $[a]$ stands for the integer part of a number $a > 0$.) Let

$$M_{ni}^{(1)} = \max\{X_j \colon X_j \in V_i\}$$

and let $M_{ni}^{(2)}$ denote the second largest element in the same group V_i. Now let us denote

$$\kappa_{ni} = \frac{M_{ni}^{(2)}}{M_{ni}^{(1)}}, \qquad S_n = \sum_{i=0}^{n} \kappa_{ni}, \qquad Z_n = n^{-1} S_n.$$

Although in [5] random variables κ_{ni} were defined and the result proved only for a sample from multivariate stable law with $0 < \alpha < 1$, it is easy to see that the main relation (6.1) from [5], used in the proof holds in the present setting and without any restriction on α, therefore the same proof gives the following result. As usual, $\xrightarrow{\text{a.s.}}$ and $\xrightarrow{\text{D}}$ denote almost sure convergence and convergence in distribution, respectively, and $N(a, \sigma^2)$ stands for the normal random variable with mean a and variance σ^2.

THEOREM A ([5]). *Let $X_1, \dots, X_N$ be a sample from a distribution satisfying* (1) *and let $n = m = [\sqrt{N}]$. Then*

$$Z_n \xrightarrow{\text{a.s.}} \frac{\alpha}{\alpha + 1}.$$

If F satisfies (2) *with $\beta = 2\alpha$, then*

$$n\big(n^{-1} S_n - \alpha(1+\alpha)^{-1}\big)\left(\sum_{i=1}^{n} \kappa_{ni}^2 - n^{-1} S_n^2\right)^{-1/2} \xrightarrow{\text{D}} N(0, 1) \tag{3}$$

and

$$n^{1/2}\big(n^{-1} S_n - \alpha(1+\alpha)^{-1}\big) \xrightarrow{\text{D}} N(0, \sigma^2) \tag{4}$$

with

$$\sigma^2 = \sigma(\alpha)^2 = \frac{\alpha}{(\alpha+1)^2(\alpha+2)}.$$

It is easy to see that the quantity $S_n(n - S_n)^{-1}$ gives a consistent and asymptotically unbiased estimator for α and it is possible for this estimator to prove the results, analogous to (3) and (4), namely

$$\frac{n(1 - Z_n)^2 (S_n(n - S_n)^{-1} - \alpha)}{\big(\sum_{i=1}^{n} \kappa_{ni}^2 - n^{-1} S_n^2\big)^{1/2}} \xrightarrow{\text{D}} N(0, 1) \tag{5}$$

and

$$n^{1/2}(S_n(n - S_n)^{-1} - \alpha) \xrightarrow{\text{D}} N(0, \tilde{\sigma}^2), \tag{6}$$

where $\tilde{\sigma}^2 = \tilde{\sigma}(\alpha)^2 = \alpha(\alpha+1)^2(\alpha+2)^{-1}$. From any of relations (3)–(6), it is possible to construct positive intervals for the unknown parameter α (in case of (4) and (6) instead of $\sigma(\alpha)^2$ and $\tilde{\sigma}(\alpha)^2$, one can use $\sigma(\hat{\alpha})^2$ and $\tilde{\sigma}(\hat{\alpha})^2$, respectively).

But it is difficult to expect a satisfactory result, since all these relations were obtained under the assumption that $n = m$, which has no justification. Therefore, now we shall work out how to choose n and m in a better way. We consider m as an independent variable, then $n = n(m) = [N/m]$, but for simplicity, we will use the relation $n = N/m$. We always assume that $n \to \infty$ and $m \to \infty$, as $N \to \infty$. Only in the case of Pareto distribution will we see that it is better to take $m = 2$ (then we use all samples). This corresponds to the well-known fact that for Pareto distribution in the Hill estimator $\gamma^{(1)}_{n,k}$, we take $k = n$. Let us denote

$$a_m = E\kappa_{n1} = \frac{\alpha}{1+\alpha} + \gamma_m, \qquad p = \frac{\alpha}{1+\alpha}, \qquad \hat{p} = n^{-1}S_n.$$

Using this notation, we can write

$$\begin{aligned}&\sqrt{n}(n^{-1}S_n - \alpha(1+\alpha)^{-1}) \\ &\quad = \sqrt{n}(\hat{p} - p) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\kappa_{ni} - E\kappa_{ni}) + \sqrt{n}\gamma_m. \end{aligned} \tag{7}$$

Now the idea is to choose n and m in such a way that the first term, which is a sum of i.i.d. bounded random variables (but forming a triangular array), converges in distribution to $N(0, \sigma^2)$, where $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$, $\sigma_n^2 = E(\kappa_{n1} - E\kappa_{n1})^2$. The second term, which is biased for the estimator $\hat{p}$, at least must stay bounded. The first possibility is to achieve $\sqrt{n}\gamma_m \to 0$, and in this way we have the following result:

THEOREM. *Let us suppose that F satisfies* (2) *with $\alpha < \beta \leqslant \infty$. If we choose*

$$n = \varepsilon_N N^{2\zeta/(1+2\zeta)}, \qquad m = \varepsilon_N^{-1} N^{1/(1+2\zeta)},$$

where $\varepsilon_N \to 0$, as $N \to \infty$ and $\zeta = (\beta - \alpha)/\alpha$, then

$$\sqrt{n}(\hat{p} - p) \xrightarrow{\mathrm{D}}_{N\to\infty} N(0, \sigma^2), \tag{8}$$

where $\sigma^2 = \lim_{n\to\infty} \sigma_n^2 = \alpha((\alpha+1)^2(\alpha+2))^{-1}$.

Another possibility is to keep the bias positive, but bounded. Using the estimate from above for γ_m, we can choose n and m in such a way that, for the first term in (7), the central limit theorem holds and

$$\sqrt{n}\gamma_m \leqslant \lambda, \tag{9}$$

with some positive λ. Then the mean-square error is

$$E\left(n^{-1}S_n - \alpha(1+\alpha)^{-1}\right)^2 = \frac{\sigma_n^2 + n\gamma_m^2}{n} \leqslant \frac{\sigma^2 + \lambda^2}{n}.$$

Since n depends on λ, we can choose λ in order to minimize this last expression. We will show that $\lambda_{\min} = \sigma(2\zeta)^{-1/2}$ and the corresponding value of n, which we believe is close to optimal, is given by formula

$$n = N^{2\zeta/(1+2\zeta)} C_0^{-2/(1+2\zeta)} \sigma^{4\zeta/(1+2\zeta)} (1+2\zeta)(2\zeta)^{-2\zeta/(1+2\zeta)},$$

where C_0 is a constant from the lemma below. Since we have only estimate (9), but not the relation $\sqrt{n}\gamma_m \to \lambda$, we can now only accertain that $\sqrt{n}(\hat{p} - p)$ is close to $N(C(\lambda_{\min}), \sigma^2)$, where $C(\lambda_{\min})$ is some bias for which we know that $0 \leqslant C(\lambda) < \lambda$.

It is not easy to compare (in the manner of [7]) this new estimator with all those listed at the beginning of the introduction, but we intend to do this in the near future. The main difficulty is caused by the fact that we have no exact value, but only an estimate for a limited bias. One more reason preventing us from comparing different estimators is the following. As was mentioned in [12], all these results concerning optimal k in the case of the Hill estimator (similar situation is for n in our case) depend on parameters α and β, which are unknown and therefore it is impossible to use in practice. Moreover, all these results have an asymptotic nature and there is no information available for moderate sample sizes. Thus these results mainly have theoretical values and show what can be expected, but for practical application, most probably the best way is to make some plot, similar to that which is used in the case of the Hill estimator (see [12]) when we plot the values $\{(k, \gamma^{(1)}_{n,k}), 1 \leqslant k \leqslant n-1\}$. Namely, we can calculate the estimator $\hat{p} = \hat{p}_m$ as a function of m, starting with small values of m. Then we plot $\{(m, \hat{p}_m), m_0 < m < M_0\}$ with some $m_0 > 2,\ M_0 < N/2$. We call such a plot a usual plot, since we shall now speak of another type of plot. In [12] it is recommended together with the Hill plot to use another plot (called the alternate Hill plot, when one plots $\{(\theta, \gamma^{(1)}_{n,[n^\theta]}),\ 0 < \theta < 1\}$), since in some situations, one plot is more informative, while in another one is more informative. We made such alternate plots for our estimator too (that is, we plotted values $\{(\theta, \hat{p}_{[N^\theta]}),\ 0 < \theta < 1\}$ and we call such a plot an alternative plot) and it seems that both plots are rather informative. From this small simulation study (we generated samples only from symmetric stable and Pareto distributions with several values of α), it is difficult to give preference to one or another type of plot, maybe countrarily, both plots, suggesting the same value of the parameter under the estimation makes in practitioner more confident. At this point, it is appropriate to mention that using statistical package S-PLUS is rather simple when making these plots and to save computer time (using a 166 Mhz Pentium processor) for moderate samples (up to $N = 10^3$) were seconds for one plot.

We generated a large number N of values of symmetrically stable or Pareto random variables with parameter α (the Pareto distribution function is defined as $F(x) = 1 - x^{-\alpha},\ x \geqslant 1$) and applied the estimator $\hat{p}$ to estimate the value of $p = \alpha/(1+\alpha)$. Only a small part of the simulation results are presented in three figures (in the preprint [18] there were six figures). In Figure 1 are the results for

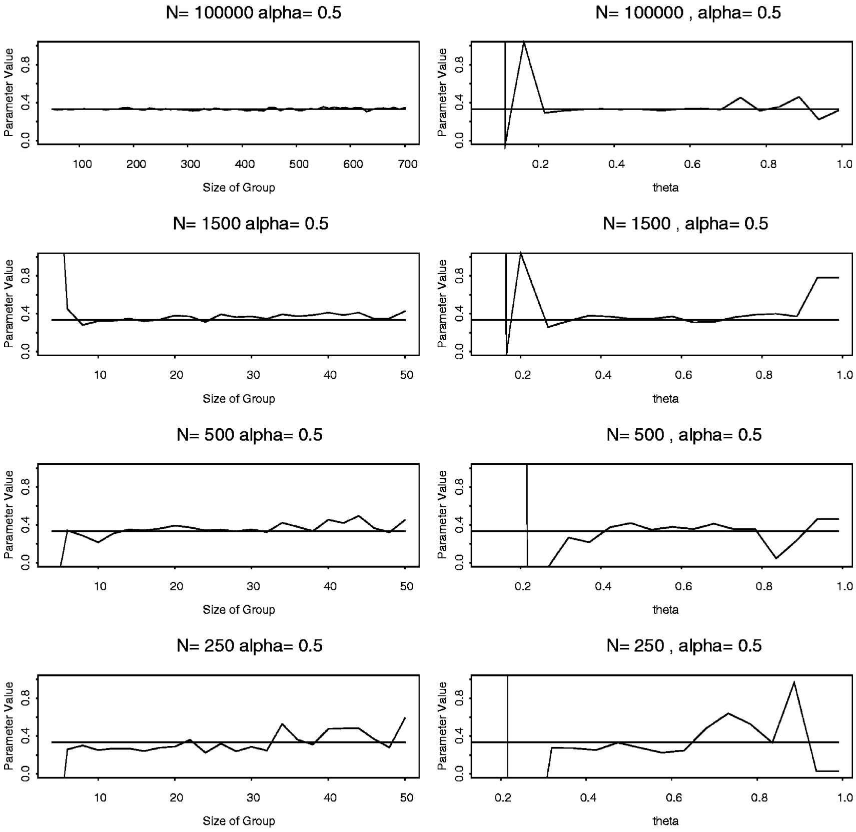

Figure 1. Stable distribution with $\alpha = 0.5$. The left column – usual plots, the right – alternative. Horizontal line – true value of p.

a symmetric stable random variable with $\alpha = 0.5$. In the left column there are that usual plots, while on the right there are alternative plots for the same values of N and α. In all the plots, the horizontal line corresponds to the true value of the parameter p. Since, for all stable random variables $\beta = 2\alpha$, $\zeta = 1$, and the asymptotically optimal value of m is approximately $N^{1/3}$ or in the alternative plot $\theta = 1/3$. For a very large sample size ($N = 10^5$, $\alpha = 1.5$) both plots are very good and there is a large range for the size of groups m and θ where the graph remains a true value of the parameter $p = \alpha/(1+\alpha) = 0.6$. For moderate values of N, ranging from 125 to 1500, it seems that the usual plots are less volatile but, as was noted by the referee, this conclusion is not true, since in the usual plot we do not show the behavior of the plot for large values of group size, while in the alternative plot we take all the values of θ, $0 < \theta < 1$. Also, it seems that the optimal value of θ is shifted to the right and is around 0.4, but this can be explained, since for smaller values of N, the role of constants is more substantial.

The situation is even better in the case of Pareto distribution (see Figure 2) with the same value of parameter $\alpha = 0.5$ as in Figure 1. For this distribution $\beta = \infty$,

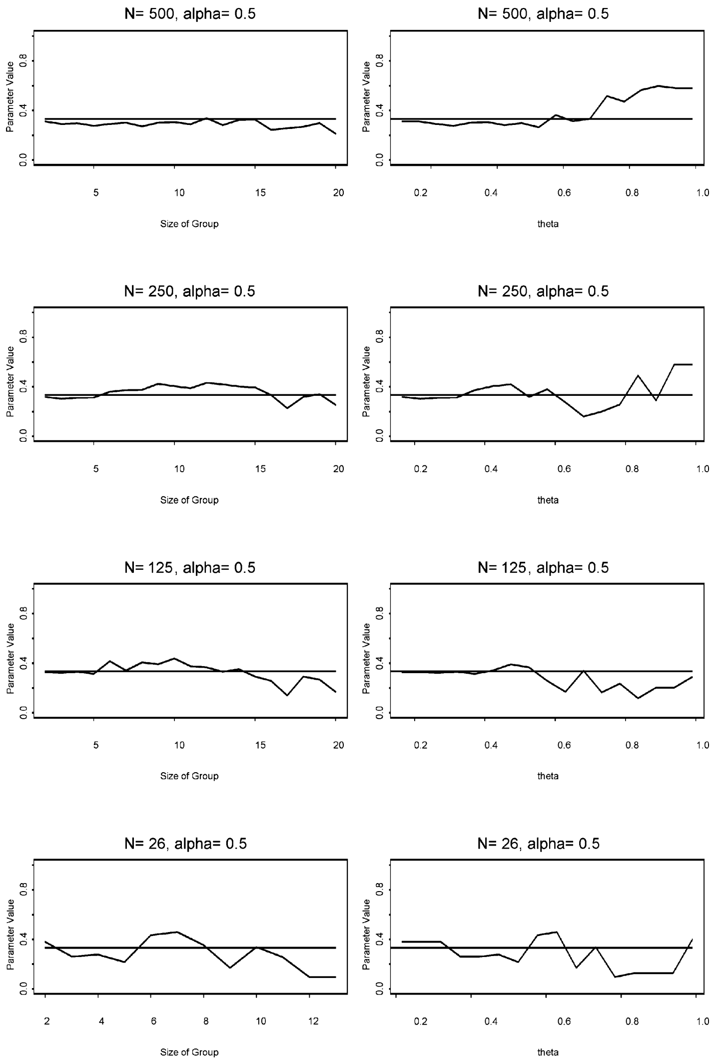

Figure 2. Pareto distribution with $\alpha = 0.5$. The left column – usual plots, the right – alternative. Horizontal line – true value of p.

therefore $\zeta = \infty$, and the best value of m is the smallest possible value, that is, 2. Even for small sample sizes N (ranging from 26 to 500) both plots – usual and alternative – for small values of m and θ give very good coincidence with true value p. Also we made simulations taking Pareto variables with bigger values of α (equal to 4 and 8), and the plots were better, reflecting the fact that the variance of the limit law for the estimator is decreasing with α increasing (see (8)). It is not difficult to verify that

$$E\left(\frac{\min(X, Y)}{\max(X, Y)}\right),$$

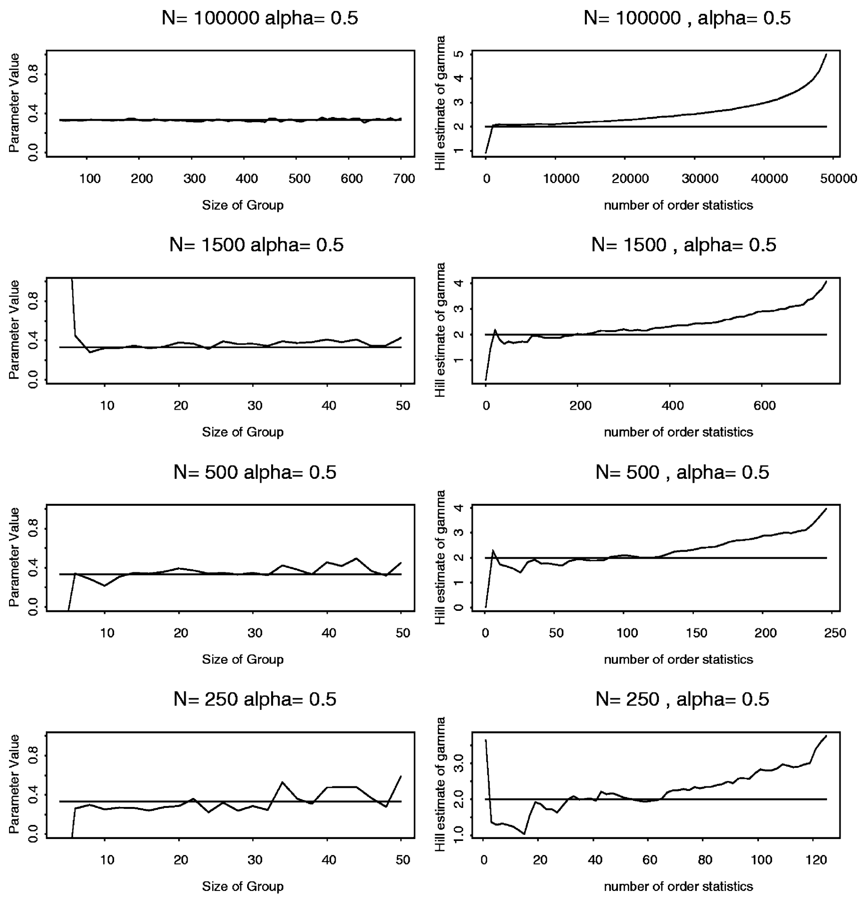

Figure 3. Stable distribution with $\alpha = 0.5$. The left column – estimation of parameter p, the right – Hill's plot of the parameter γ. Horizontal line – true value of parameters.

where X, Y are two independent Pareto random variables with exponent α, is exactly $p = \alpha(1+\alpha)^{-1}$, therefore our estimator $\hat{p}$ is nothing but the sample mean for a bounded random variable.

Figure 3 compares our estimator $\hat{p}$ with Hill's estimator $\gamma_{n,k}^{(1)}$. We took the same samples of stable (in [18] there is also a comparison with Pareto distribution) with $\alpha = 0.5$, used in Figure 1, thus the left column of Figure 3 is simply the repeated left column of Figure 1. On the right column, there are Hill plots of the estimator $\gamma_{n,k}^{(1)}$, which estimates parameter $\gamma = 1/\alpha$, so in these graphs, the horizontal line is at $\gamma = 1/0.5 = 2$.

All these figures suggest that our new estimator's performance is quite good, and it can be explained (good performance in the Pareto case as we just explained) that despite the fact that it uses not all the largest values from the sample, it has a very simple form – it is a sum of i.i.d. bounded random variables. Therefore, it seems that it is possible to recommend that practitioners try this new estimator on real data. At the same time, it is necessary to note that some problems remains to be investigated. While this estimator is robust to scale transformation (the value of $\hat{p}$ remains unchanged if we multiply all samples by some positive constant), it is sensitive to a shift. Also, it remains an open question just how this estimator behaves when we drop the assumption of independence in the sample. One more direction of investigation is to use time which the plot spends in the vicinity of the horizontal line, corresponding to the true value of the parameter under consideration. This idea was discussed during the conference in the talk by H. Drees. Preliminary simulations show that plots of our new estimator are well suited to exploit this idea. We intend to investigate these questions in near future.

2. Proof of the Theorem

As was explained before in the formulation of the theorem, in order to prove (8), we need to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\kappa_{ni} - E\kappa_{ni}) \to N(0, \sigma^2) \tag{10}$$

and

$$\sqrt{n}\gamma_m \to 0.$$

Since (10) easily follows (see [5]), the main step in the proof is to estimate γ_m. If we denote $G(x) = 1 - F(x)$, then κ_{n1} has the same distribution as

$$G^{-1}\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)\left(G^{-1}\left(\frac{\Gamma_1}{\Gamma_{m+1}}\right)\right)^{-1},$$

where G^{-1} is the inverse function for G and $\Gamma_i = \sum_{j=1}^{i}\lambda_i$ with $\lambda_i,\ i \geqslant 1$, being i.i.d. standard exponential random variables. Therefore

$$\gamma_m = E\left(G^{-1}\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)\left(G^{-1}\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)\right)^{-1}\right) - \frac{\alpha}{1+\alpha}.$$

LEMMA. *Suppose that F satisfies* (2) *with $\alpha < \beta \leqslant \infty$. Then*

$$|\gamma_m| \leqslant C_0 m^{-\zeta}, \tag{11}$$

where C_0 is a constant depending on $C_1, C_2, \alpha,$ and β.

Proof. We have that, for large values of x,

$$G(x) = C_1 x^{-\alpha} + C_2 x^{-\beta} + \mathrm{o}(x^{-\beta}),$$

therefore it is possible to show that for the inverse function for small values of t, we have the following relation:

$$G^{-1}(t) = at^{-1/\alpha} + bt^{-1/\alpha+(\beta-\alpha)\alpha^{-1}} + \mathrm{O}\left(t^{-1/\alpha+2(\beta-\alpha)\alpha^{-1}}\right) \tag{12}$$

with $a = C_1^{1/\alpha},\ b = \alpha^{-1} C_2 C_1^{(1-\beta)\alpha^{-1}}$. From (12) it follows that for small $\delta > 0$ and $0 < t < \delta$,

$$G^{-1}(t) t^{1/\alpha} C_1^{-1/\alpha} = 1 + C_1^{-1/\alpha} b t^{\zeta} \bigl(1 + \mathrm{O}(t^{\zeta})\bigr).$$

Therefore, if we take δ such that $|\mathrm{O}(t^{\zeta})| \leqslant 1/2$ (this gives us $\delta = \mathrm{o}(2^{-1/\zeta})$), then we can write

$$1 + C_4 t^{\zeta} \leqslant G^{-1}(t) t^{1/\alpha} C_1^{-1/\alpha} \leqslant 1 + C_3 t^{\zeta} \tag{13}$$

with $C_3 = (3/(2\alpha)) C_2 C_1^{-\beta/\alpha},\ C_4 = \frac{1}{3} C_3$.

As in [5], we use the relation

$$E\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{1/\alpha} = \frac{\alpha}{\alpha + 1}.$$

Let

$$R_+^{m+1} = \{\bar{x} = (x_1, \ldots, x_{m+1}) : x_i \geqslant 0, i = 1, \ldots, m+1\},$$
$$\Sigma_m = x_1 + \cdots + x_m;$$
$$[A_{2,m} = \{\bar{x} \in R_+^{m+1} : (x_1 + x_2)/\Sigma_{m+1} \geqslant \delta\}, \qquad A_{2,m}^c = R_+^{m+1} \backslash A_{2,m},$$

where δ is such that (13) holds. Then

$$|\gamma_m| = \left| E\left(G^{-1}\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)\left(G^{-1}\left(\frac{\Gamma_1}{\Gamma_{m+1}}\right)\right)^{-1}\right) - \frac{\alpha}{\alpha+1} \right| \leqslant I_1 + I_2 + I_3, \tag{14}$$

where

$$I_1 = \int_{A_{2,m}} \left| G^{-1}\left(\frac{x_1 + x_2}{\Sigma_{m+1}}\right)\left(G^{-1}\left(\frac{x_1}{\Sigma_{m+1}}\right)\right)^{-1} \mathrm{e}^{-\Sigma_{m+1}} \mathrm{d}\bar{x} \right|,$$
$$I_2 = \int_{A_{2,m}}^{c} \left| G^{-1}\left(\frac{x_1 + x_2}{\Sigma_{m+1}}\right)\left(G^{-1}\left(\frac{x_1}{\Sigma_{m+1}}\right)\right)^{-1} - \left(\frac{x_1}{x_1 + x_2}\right)^{1/\alpha} \right| \mathrm{e}^{-\Sigma_{m+1}} \mathrm{d}\bar{x},$$

and

$$I_3 = \int_{A_{2,m}} \left(\frac{x_1}{x_1 + x_2}\right)^{1/\alpha} \mathrm{e}^{-\Sigma_{m+1}} \mathrm{d}\bar{x}.$$

Since the integrands in both integrals I_1 and I_3 are positive and less than 1, we easily get for any $\tau > 0$ (which will be chosen later)

$$I_1 + I_3 \leqslant 2P\left\{\frac{\lambda_1 + \lambda_2}{\lambda_1 + \cdots + \lambda_{m+1}} > \delta\right\} \leqslant 2\delta^{-\tau} E\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)^{\tau}. \tag{15}$$

If we denote

$$t_1 = \frac{x_1}{\Sigma_{m+1}}, \qquad t_2 = \frac{x_1 + x_2}{\Sigma_{m+1}}, \qquad U = \frac{G^{-1}(t_2)}{G^{-1}(t_1)}, \qquad A = \left(\frac{t_1}{t_2}\right)^{1/\alpha},$$

then we have $t_1 < t_2 < \delta$ in $A^c_{2,m}$, therefore, applying (13), we can write

$$AB_1 \leqslant U \leqslant AB_2,$$

where

$$B_1 = \frac{1 + C_4 t_2^\zeta}{1 + C_3 t_1^\zeta}, \qquad B_2 = \frac{1 + C_3 t_2^\zeta}{1 + C_4 t_1^\zeta}.$$

Then

$$|U - A| \leqslant A \max((B_2 - 1), |B_1 - 1|). \tag{16}$$

Taking into account that $C_3 = 3C_4$ and applying rough estimates, we get

$$\max((B_2 - 1), |B_1 - 1|) \leqslant 4C_4 t_2^\zeta. \tag{17}$$

From (16) and (17) we have

$$\begin{aligned} I_2 &= \int_{A^c_{2,m}} |U - A| \exp(-\Sigma_{m+1})\, \mathrm{d}\bar{x} \\ &\leqslant 4C_4 \int_{A^c_{2,m}} A t_2^\zeta \exp(-\Sigma_{m+1})\, \mathrm{d}\bar{x} \leqslant 4C_4 E\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)^\zeta. \end{aligned} \tag{18}$$

It remains to evaluate $E(\Gamma_2/\Gamma_{m+1})^\zeta$. It is known (see, for example, [1]) that m-dimensional random vector $(\Gamma_1/\Gamma_{m+1}, \ldots, \Gamma_m/\Gamma_{m+1})$ has the density

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} n!, & \text{if } 0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n \leqslant 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Then it is not difficult to find that Γ_2/Γ_{m+1} has the density (we assume that $m \geqslant 2$)

$$g(x) = \begin{cases} m(m-1)x(1-x)^{m-2}, & \text{if } 0 \leqslant x \leqslant 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Therefore

$$\begin{aligned} E\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)^\zeta &= m(m-1)\int_0^1 x^{1+\zeta}(1-x)^{m-2}\, \mathrm{d}x \\ &= m(m-1)B(m-1, \zeta + 2), \end{aligned} \tag{19}$$

where $B(\mu, \nu) = \int_0^1 x^{\nu-1}(1-x)^{\mu-1}\, \mathrm{d}x$ is the so-called Beta-function. From tables (see, for example, [17]) one can get the following relation between this function and well-known Gamma-function:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}.$$

Performing simple calculations, we have

$$B(m-1, \zeta+2) \leqslant 4\Gamma(\zeta+2)m^{-(2+\zeta)},$$

therefore

$$E\left(\frac{\Gamma_2}{\Gamma_{m+1}}\right)^{\zeta} \leqslant 4\Gamma(\zeta+2)m^{-\zeta}. \tag{20}$$

Collecting estimates (14), (15), (18)–(20) and choosing $\tau = \zeta$ in (15), we get the estimate (11) with $C_0 = C_0(\alpha, \beta, C_1, C_2)$ (it is possible to write down this constant, but since we are not interested in exact values of constants, we will not do this). The lemma is proved. □

Having the estimate for γ_m, we choose a sequence $\varepsilon_N \to 0$, as $N \to 0$, and taking

$$n = \varepsilon_N^{2/(1+2\zeta)} C_0^{-2/(1+2\zeta)} N^{2\zeta/(1+2\zeta)},$$
$$m = \varepsilon_N^{-2/(1+2\zeta)} C_0^{2/(1+2\zeta)} N^{1/(1+2\zeta)}, \tag{21}$$

we get $\sqrt{n}\gamma_m \to 0$. Therefore, taking into account (6), we get (8) and the theorem is proved. □

If for some $\lambda > 0$, we chose n and m from the relation

$$C_0\sqrt{n}m^{-\zeta} = \lambda,$$

namely, if n and m are given by (21) only with λ instead of ε_N, then, as was explained after the formulation of the theorem, the mean square error $E((1/n)S_n - (\alpha/(1+\alpha)))^2$ is approximately equal to $1/n(\lambda^2+\sigma^2)$. Since n depends on λ, we can choose λ in order to minimize this expression. It is easy to see that this minimizing value is $\lambda_{\min} = \sigma/\sqrt{2\zeta}$ and the corresponding value of n is given by the following formula:

$$n_{\text{opt}} = \sigma^{4/(1+2\zeta)} \tilde{C}_0^{-2/(1+2\zeta)} N^{2\zeta/(1+2\zeta)},$$

where $\tilde{C}_0 = C_0(2\zeta)^{\zeta}(1+2\zeta)^{-(1+2\zeta)/2}$.

Acknowledgement

The author is grateful to Ronald Shonkwiler and Remigijus Lapinskas who helped in dealing with S-PLUS.

References

1. Breiman, L.: *Probability*, Addison-Wesley, Reading, Mass., 1968.
2. Beirlant, J., Teugels, J. L. and Vynckier, P.: Tail index estimation, Pareto quantile plots, and regression diagnostics, *J. Amer. Statist. Assoc.* **91** (1996), 1659–1667.

3. Csörgő, S., Deheuvels, P. and Mason, D.: Kernel estimates of the tail index of a distribution, *Ann. Statist.* **13** (1985), 1050–1077.
4. Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G.: Using a bootstrap method to choose the sample fraction in tail index estimation, *J. Multivariate Anal.* **76** (2001), 226–248.
5. Davydov, Yu., Paulauskas, V. and Račkauskas, A.: More on P-stable convex sets in Banach spaces, *J. Theoret. Probab.* **13**(1) (2000), 39–64.
6. Davydov, Yu. and Paulauskas, V.: On the estimation of the parameters of multivariate stable distributions, *Acta Appl. Math.* **58** (1999), 107–124.
7. de Haan, L. and Peng, L.: Comparison of tail index estimators, *Statist. Nederlandica* **52**(1) (1998), 60–70.
8. de Haan, L. and Resnick, S.: Method to choose the sample fraction in tail index estimation, *J. Roy. Statitst. Soc. Ser. B* **42** (1980), 83–87.
9. Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L.: A moment estimator for the index of an extreme-value distribution, *Ann. Statist.* **17** (1975), 1833–1855.
10. Dekkers, A. L. M. and de Haan, L.: Optimal choice of sample fraction in extreme-value estimation, *J. Multivariate Anal.* **47**(2) (1993), 172–173.
11. Draisma, G., de Haan, L. and Peng, L.: A Bootstrap-based method to achieve optimality in estimating the extreme-value index, *Extremes* **2**(4) (1999), 367–404.
12. Drees, H., de Haan, L. and Resnick, S.: How to make a Hill plot, *Ann. Statist.* **28**(1) (2000), 254–274.
13. Drees, H. and Kaufmann, E.: Selecting the optimal sample fraction in univariate extreme value estimation, *Stochastic Process. Appl.* **75**(2) (1998), 149–172.
14. Hall, P.: Using the bootstrap to estimate means squared error and select smoothing parameter in nonparametric problems, *J. Multivariate Anal.* **32** (1990), 177–203.
15. Hill, B. M.: A simple general approach to inference about the tail of a distribution, *Ann. Statist.* **3** (1975), 1163–1174.
16. Geluk, J. L. and Peng, L.: An adaptive optimal estimate of the tail index for MA(1) time series, *Statist. Probab. Lett.* **46** (2000), 217–227.
17. Korn, G. A. and Korn, T. M.: *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, New York, 1961.
18. Paulauskas, V.: A new estimator for tail index, Preprint, Vilnius University, 2002.
19. Pickands, J.: Statistical inference using extreme order statistics, *Ann. Statist.* **3**(1) (1975), 119–131.